公開

（事後評価・研究者用）

研　究　報　告　書

「大規模テキストからの知識獲得と深層学習による照応・省略解析」

研究期間：平成 29 年 10 月～平成 31 年 3 月
研究者番号：５０１３７
研 究 者：栗田　修平

1．　研究のねらい

人間のように文章を理解するシステムの開発には、日常会話で用いられるようなこなれた文や文章を理解することが欠かせない。そのような文の中では、常識的な事項や、文章の著者および読者にとって既知と思われる事項が省略される傾向がある。このような文を解析するためには、自然言語処理の中でも、既知の事項をその文脈および類似した文中での用例から補う省略解析や、文のもつ意味を解析する意味解析が欠かせない処理となる。特に、省略解析は、自然な文章中に多く出現する省略される対象を推定するタスクで、人間が日常生活の中で接するこなれた会話や文章などを処理するためには必須の操作である。一方で、その精度は大きく伸び悩んでおり、人間との円滑なコミュニケーションを目指すシステムに自然言語処理を応用する際、大きな障害となってきた。本研究にて実現される省略解析およびその統合モデルを使用することで、より人間に近い対話や翻訳システムが実現可能となる。具体的には、上に挙げた省略解析の統合モデルの提案・実証後に具体的な応用として、可能な限り人間の意図を汲み取って会話が行える対話システム、会話に特化した翻訳エンジンを使用したリアルタイム多言語チャットサービスなどが考えられる。また、日常会話に基づくスマートフォンやカーナビゲーションなどの操作サービスや、柔軟な会話が可能な対話型の接客・注文サービスなど、様々な分野での新しい応用が期待される。

本研究では、近年、急速な発展を遂げつつある深層学習手法を利用して、大規模テキストから獲得された知識および前後の文脈を考慮した省略解析手法を提案し、その精度を既存手法とは一線を画するほどまで引き上げることを目指す。また、単一タスクでの学習では限界があった、より複雑な知識処理の統合を模索する。まず、Web ページから収集された膨大なテキストを処理して大規模な学習データを作成し、省略解析に必要な常識知を学習させる。加えて、文章の文脈を考慮させるため、近年大きな発展を遂げている深層強化学習の文の意味解析への応用を模索する。このようにして、既存手法ではなかなか困難とされてきた省略解析のさらなる高度化を目指す。

2．　研究成果

（1）概要

本研究では、自然言語処理において文内での単語の意味的な関係や省略された事項を解析するための新たなモデルを提案し、特に深層学習の中でも生成的なアプローチおよび深層強化学習と組み合わせることで、既存手法における限界を超える性能を達成するモデルを提案した。本研究では、文の中の多様な意味に対する表現を解析するため、まずは(1) 日本語の述語項構造解析に基づく省略解析モデルに対する敵対的生成ネットワーク(GAN)による学習法を応用した半教師付き学習法を提案し、次に、(2) 英語の意味依存構造解析を深層強化学習により行うアルゴリズム及びモデルを提案した。日本語の省略解析につい

ては、平易で日常的な文に対する解析ほど難しいことが知られている。例えば「空は曇っているけれど、いまから外に干しても大丈夫？」というような例文について、「干して」のヲ格に相当する部分はこの文中に示されていない。このように自然な文章の中で省略されている事項は、主にその周辺の文脈に幅広く依存し、外部知識と呼ばれる一般的な知識も必要となる。このように文脈や外部知識に依存する問題の解決のために、多様な文脈の中で省略された表現を推測する学習データを用意する必要がある。そこで近年活発に研究が行われている敵対的生成ネットワークを応用し、大規模コーパスの処理による学習データの生成とモデルの学習とを両立する手法を提案した。さらに、英語においても深層強化学習手法を利用して文中の平易な箇所から処理を行う新規なモデルを提案した。英語の単語間の係り受けは、従来型の依存係り受け解析のような構造に近いものから、日本語の述語項構造解析に近いものまでが存在する。本研究では、深層強化学習を利用して平易な箇所から順番にグラフを構築していくアルゴリズムを提案し、さらに提案モデルを用いて実証した。今後は、これらの学習手法を深層の事前学習手法や転移学習と組み合わせる予定である。

これらの研究成果のうち、日本語の省略解析については、自然言語処理のトップカンファレンスである ACL2018 に採択され発表を行うなど国際的に高く評価されている。また、深層強化学習を用いた意味依存構造解析は言語処理学会第 25 回年次大会にて発表し最優秀賞に選ばれるなど高く評価されつつある。（主な研究成果リスト[2],[3],[4]）

（2）詳細

本研究では、(1) 敵対的生成ネットワークを利用した日本語の省略解析 (2) 深層強化学習による英語意味依存構造解析の 2 つの分野について、大きな成果を達成した。

(1) 敵対的生成ネットワークを利用した日本語の省略解析

日本語の省略解析については、自然な文章の中で省略されている事項を、その文脈から推測するモデルを提案した。文章中で省略されている事項を推測するためには、多様な文脈の中で省略された表現を推測する学習データを用意する必要がある。しかし、省略解析に用いられる日本語述語項構造解析データセットは高々数千文相当しか存在せず、多様な文脈の中で学習できるとは言い難い。この学習データの不足を補うために、近年盛んに研究が行われている敵対的生成ネットワークを応用した新しいニューラルネットワークモデルを提案した。具体的には、Web から獲得された大規模テキストを生成ニューラルネットワークおよび分類ニューラルネットワークの 2 つのニューラルネットワークで処理を行い、データセットの生成と解析の双方を同時に行うニューラルネットワークを作成した。また、特に日常的に使われやすい表現を多数含む京都大学リードコーパスでの例文に焦点を当てて解析を行なった。これは、WEB ページから作成されたリードコーパスには、日常表現が多くの文中に出現する一方で、このような文を解析することは未だ難しく、社会的な需要も大きいものと考えられるからである。本研究は、自然言語処理のトップカンファレンスである ACL2018 にて、long paper として採択され、発表を行った。（主な研究成果リスト[2]）

(2) 深層強化学習による英語意味依存構造解析

英語において文内の単語同士の意味的な関係を解決する意味依存構造解析の新規なモデルを提案した。英語の単語間の関係は、従来は主として依存係り受け解析のような文法的

な解析により解決されてきていた。しかし、こうした解析には単語間の関係の表現に制約が大きく、より柔軟に文内の意味を表現できているとは言いがたかった。本研究では、文の中の単語同士の複雑な関係をより柔軟に表現できる意味依存構造解析に焦点を当てた。意味依存構造解析では、単語間の意味的な関係性は係り受け解析よりも複雑なグラフで表せられる。本研究では、そのようなグラフに対し深層強化学習を利用して、平易な箇所から順番に構築していくモデルを提案した。本研究成果は、言語処理学会第 25 回年次大会にて発表し、最優秀賞に選ばれた。(主な研究成果リスト[3])
これらの実験はその多くが GPU 計算機上で行われ、本研究では GPU 計算基盤の整備および GPU 計算技術の精緻化をも同時に追求した。

3. 今後の展開

自然言語文の高度かつ高精度な解析は、自然言語処理が社会に広まる際には必須となる基盤的な技術であり、その応用範囲は、機械翻訳モデルや対話エンジン、SNS などの高度な感情分析など非常に多岐にわたる。日本語の解析については、本研究にて行った省略解析は今後の自然言語処理の基盤となる解析である。英語の意味依存構造解析についても、同様に、既存の解析においては深く扱われてこなかった単語内の意味的な関係について、従来法とは異なるアプローチによる解決を行っている。いずれの研究も、今後はより一般的な知識という観点から、事前学習及び転移学習と組み合わせた研究として発展させていく予定である。また、得られたモデルを一般ユーザが利用できる形での公開を検討している。

4. 自己評価

研究目的の達成状況については、 深層学習を用いた日本語の省略解析に加えて、英語の意味依存構造解析も行うなど、当初の予定を大きく超えて進展していると思われる。このような成果に加えて、転移学習や社会的な応用研究などの点においての進展が望まれる。研究の進め方や研究体制については、概ね計画のとおり進展していると考えられる。なお、コペンハーゲン大学 A. Søgaard と深層強化学習を用いた意味依存構造解析に対する共同研究を行った。研究成果の波及効果については、深層学習を利用した文の解析手法により、既存手法では不可能であった解析が行えるようになった。この研究成果は、自然言語処理を利用した機械翻訳や対話ボットなどの応用が幅広い期待される。ACT-I 加速フェーズにおいては、こうした成果物の API やモデルとしての公開の他に、一般ユーザーが利用できる形での成果物の公開を目指している。研究課題の独創性・挑戦性については、本研究課題は、深層学習研究の中でも先進的な敵対的生成ネットワーク(GAN)や深層強化学習を自然言語文の解析に応用した、非常に独創性及び挑戦性の強い研究課題に対する解説方策であったと考えている。今後は、この解決方策を発展させ、同時に事前学習や転移学習と組み合わせ成熟させることが重要であると思われる。

5. 主な研究成果リスト

(1) 論文 (原著論文) 発表

1. 栗田修平, 河原大輔, 黒橋禎夫. ニューラルネットワークを利用した中国語の統合

的な構文解析. 自然言語処理. 2019. Vol.26. 231-258.
2. Shuhei Kurita, Daisuke Kawahara and Sadao Kurohashi. Neural Adversarial Training for Semi-supervised Japanese Predicate-argument Structure Analysis. The 56th Annual Meeting of the Association for Computational Linguistics (ACL2018). 2018. P18-1044. 474-484

(2) 特許出願

研究期間累積件数:0 件

(3) その他の成果 (主要な学会発表、受賞、著作物、プレスリリース等)

[3] 栗田 修平, Anders Søgaard. 深層強化学習を用いた意味依存構造解析は自発的に平易優先戦略を学習する. 言語処理学会第 25 回年次大会, 2019. 第 25 回年次大会最優秀賞 受賞.

[4] 栗田 修平. 自然言語処理の既存データセットの制約を超えた文の解析手法. RIKEN AIP Public. 2019 年 2 月. 招待講演(国内).